Answers to the review sheet

1. a 2. e 3. c 4. b 5. d 6. E

7. a)

American League (AL)		National League (NL)
	2	9
5	3	1
930	4	26788
88741	5	3555
884	6	337
75	7	

5|3 represents an american league team which hit 35 home runs during the 1994 season

2|9 represents a national league team which hit 29 home runs during the 1994 season

American league
5 number summary
35 49 57.5 68 77

National league
5 number summary
26 46 50.5 55 67

$\bar{x}=56.93$
$s=12.69$

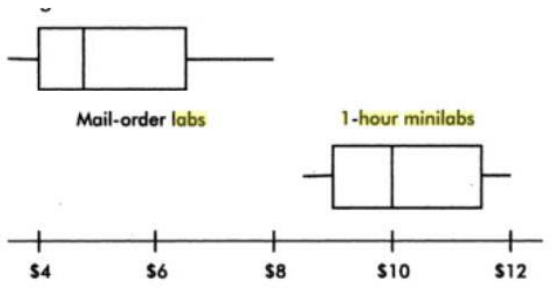
$\bar{x}=50.14$
$s=11.13$

Overall, the American League teams hit more home runs than National League teams in 1994. To illustrate this, the median for the AL was 57.5 while the median for the NL was only 50.5. Additionally, the low value for the NL is much lower than in the AL (29 versus 35) and the high value in the AL is much higher than that in the NL (77 versus 67).

8. There are other acceptable answers for this question

a. 3, 4, 4, 4, 5
b. 3, 4, 4, 4, 15
c. 1, 4, 4, 4, 5
d. 1, 2, 3, 4, 4
e. -2, -1, 0, 1, 2

9.

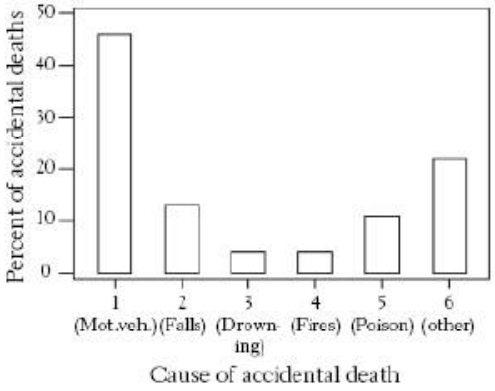


Boxplots show the minimum, maximum, median, and quartile values. The distribution of mail-order lab prices is lower and more spread out than that of prices of 1-hour minilabs. Both are slightly skewed toward the upper end (the skewness can also be noted from the computer output showing the mean to be greater than the median in both cases.)

Answers to packet questions:

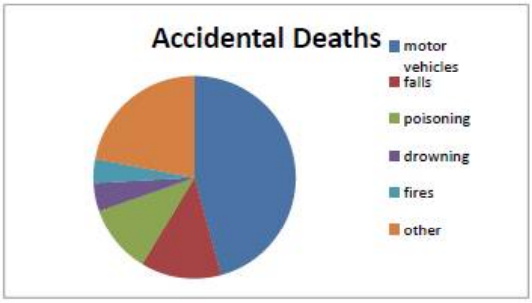
1.6 **ACCIDENTAL DEATHS** In 1997 there were 92,353 deaths from accidents in the United States. Among these were 42,340 deaths from motor vehicle accidents, 11,858 from falls, 10,163 from poisoning, 4051 from drowning, and 3601 from fires.
(a) Find the percent of accidental deaths from each of the causes, rounded to the nearest percent. What percent of deaths were due to other causes? (b) Make a well-labeled bar graph of the distribution of causes of accidental deaths. Be sure to include an "other causes" bar.

Motor Vehicles = 46 %
Falls = 13 %
Drowning = 4 %
Fires = 4 %
Poisoning = 11 %
Other causes = 22 %

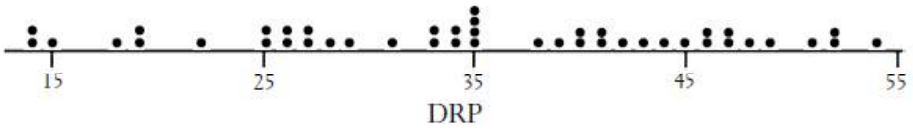


(c) Would it also be correct to use a pie-chart to display these data? If so, construct the pie-chart. If not, explain why not.

A pie chart could also be used, since the categories represent parts of a whole (all accidental deaths).



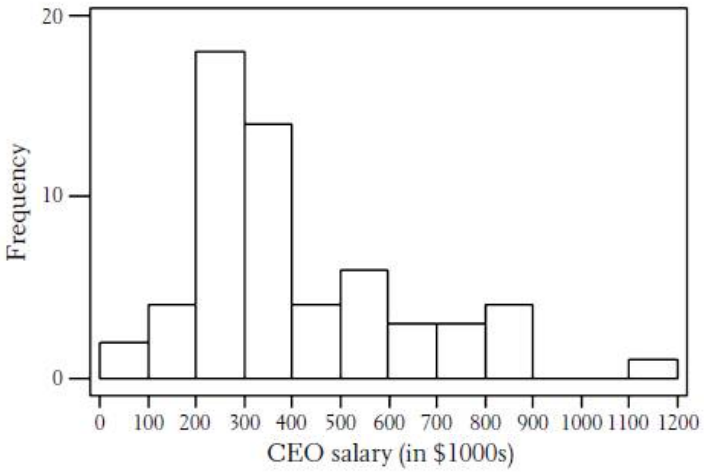
1.10



The center of the distribution is 35, and there are approximately the same number of points to the left and right of the center. There are no major gaps or outliers. The distribution is approximately symmetric.

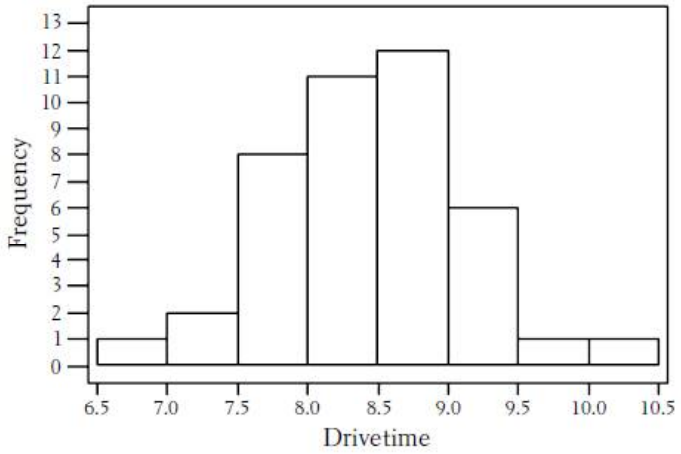
A histogram or a stem and leaf plot would also have been good ways of representing the data for 1.10

1.14



The distribution is skewed to the right with a peak in the 200s class. The spread is approximately 1100 ($21,000 to $1,103,000) and the center is located at 350 ($350,000). There is one outlier in the 1100s class.

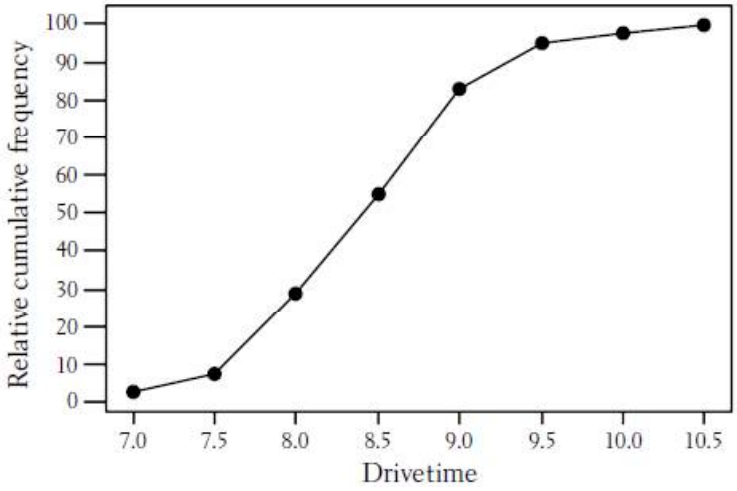
1.29 (a)



The distribution is roughly symmetric with no clear outliers.

(b)

Drivetime	Cum. freq.	Rel. cum. freq.
7.0	1	2.4%
7.5	3	7.1%
8.0	12	28.6%
8.5	23	54.8%
9.0	35	83.3%
9.5	40	95.2%
10.0	41	97.6%
10.5	42	100%



(c) Center ≈ 8.5, 90th percentile ≈ 9.4

(d) 8.0 ≈ 28th percentile

1.32 (a)

```
10 | 139
11 | 5
12 | 669
13 | 77
14 | 08
15 | 244
16 | 55
17 | 8
18 |
19 |
20 | 0
```

200 is a potential outlier. The center is approximately 140. The spread (excluding 200) is 77.

(b) $\bar{x} = 2539/18 = 141.058$.

(c) Median = average of ninth and tenth scores = 138.5. The mean is larger than the median because of the outlier at 200, which pulls the mean towards the long right tail of the distribution.

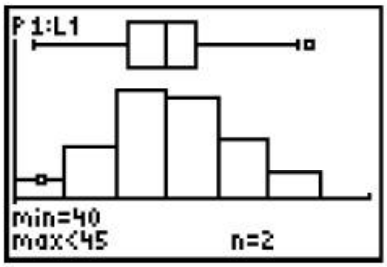
1.37 (a) The mean and median should be approximately equal since the distribution is roughly symmetric.

(b) Five-number summary: 42, 51, 55, 58, 69 $\quad \bar{x} = 2357/43 = 54.8$

As expected, median and $\bar{x}$ are very similar.

(c) Between Q_1 and Q_3: 51 to 58.

(e)



The point 69 is an outlier; this is Ronald Reagan's age on inauguration day. W. H. Harrison was 68, but that is not an outlier according to the 1.5(IQR) test.

1.41 (a) $\sum(x_i - \bar{x})^2 = (-12.1)^2 + (1.9)^2 + (-10.1)^2 + (12.9)^2 + (34.9)^2 + (6.9)^2 + (-3.1)^2 + (-.1)^2 + (-18.1)^2 + (-13.1)^2 = 2192.9$.

$s^2 = 2192.9/9 = 243.66$, $s = 15.609$.

(b) Excluding the outlier at 61, we obtain $\bar{x} = 22.2$, $s = 10.244$. The outlier caused the values of both measures to increase; the increase in s is more substantial. Clearly, s is not a resistant measure of spread.

1.44 (a) $\bar{x} = 7.5/5 = 1.5$, $s = .436$.

(b) To obtain $\bar{x}$ and s in centimeters, multiply the results in inches by 2.54: $\bar{x} = 3.81$ cm, $s = 1.107$ cm.

(c) The average cockroach length can be estimated as the mean length of the five sampled cockroaches: that is, 1.5 inches. This is, however, a questionable estimate, because the sample is so small.

1.54 The means and standard deviations are basically the same. For Set A, $\bar{x} \approx 7.501$ and $s \approx 2.032$, while for Set B, $\bar{x} \approx 7.501$ and $s \approx 2.031$. Set A is skewed to the left, while Set B has a high outlier.

Set A	
3	1
4	7
5	
6	1
7	2
8	1177
9	112

Set B	
5	257
6	58
7	079
8	48
9	
10	
11	
12	5